Vowel Duration and Spectral Balance in Scottish English and Russian

Olga Gordeeva, Ineke Mennen and James M. Scobbie
Queen Margaret University College, Edinburgh, U.K.
E-mail: ogordeeva@qmuc.ac.uk, imennen@qmuc.ac.uk, jscobbie@qmuc.ac.uk

ABSTRACT

This is a cross-linguistic study into the effect of extrinsic vowel duration conditioning on the realisation of spectral balance (intensity distribution above 0.5 kHz) in Scottish English and Russian. The materials consisted of monosyllabic words containing the vowel /i/ in prominent syllables, with varied right consonantal context in order to trigger extrinsic vowel duration conditioning. Results show, that Scottish females have a higher spectral level (between 2.5–4.5 kHz) in the short /i/ than in the long one, whereas Russian females do not differentiate between the two in spectral balance, and not to the same extent in duration. We present the results and discuss the observed differences in duration and spectral balance with regards to possible confounding effects.

1. INTRODUCTION

The aim of this study is to establish the extent of the differences in duration and spectral balance between Modern Standard Russian (Russian) and Scottish Standard English (Scottish). The study forms part of a larger study into the production of suprasegmental acoustic properties in prominent syllables by Russian/Scottish bilingual children. Adult female speech embedded in monosyllabic words (part of child vocabulary) form control for the child-language study. Due to space limitations, we will not report on fundamental frequency, but only on duration and spectral balance.

There have been no cross-linguistic studies comparing duration and intensity distribution in prominent syllables in Russian and Scottish. To our knowledge, there are no studies on the acoustic cues to prominence (other than duration) in Scottish. We have reason to believe that Russian and Scottish may differ in the use of intensity, because Scottish employs a considerable extrinsic vowel duration conditioning by the right consonantal context, known as the Scottish vowel length rule (SVLR) [1][2]. The SVLR in Scottish applies in particular to high vowels /i/ and /ʉ/: i.e. they undergo 80 to 100% lengthening, when followed by voiced fricatives (as compared to unvoiced plosives). As opposed to Scottish, Russian employs only a limited extent of vowel duration conditioning (e.g. Chen reports 20% increase for /i/ [3]). In Russian, duration is reported to be the most robust acoustic and perceptive cue to word stress and prominence, and overall intensity is considered to be a relatively weak cue [4][5].

For our study, we would like to establish: (1) the extent of the influence of the right consonantal context between the two languages on the duration of vowel /i/ in prominent syllables; (2) whether spectral balance within prominent syllables is employed to a different extent between the two languages; (3) whether the extrinsic phonetic load on the vowel duration in Scottish affects the employment of spectral balance within prominent syllables.

1.1 SPECTRAL BALANCE AND OVERALL INTENSITY

Overall intensity has proven to be only a weak perceptive cue in utterance prominent positions for both Russian [5] and American English [7][6]. Although there is a clear relationship between overall intensity and loudness, loudness is known to correlate with other factors such as segment duration, spectral balance and pitch. In fact, in a prominent – non-prominent paradigmatic comparison of syllables, the total amplitude (temporal summation of waveform amplitude through the syllable nucleus) [6]; and spectral balance (spectral levels at frequencies between 0.5 – 4 kHz)[8][9] are found to be acoustically more robust and perceptively more relevant cues to prominence than the overall intensity. The total amplitude emphasises the integration of energy of a sound in time, whereas an increase in spectral balance in high frequencies (a consequence of a more asymmetrical glottal pulse, e.g. [10][11]) is hypothesised to correlate with an increased physiological effort [8].

Linguistic stress and prominence, and their acoustic manifestations are relative properties. If acoustic properties, like duration or spectral balance, are robust cues for the prominent – non-prominent distinction in syllables, they should also be robust cues across prominent syllables (within a language and cross-linguistically). A comparison across prominent syllables is very useful in a cross-linguistic study like this, because it allows us to abstract away from cross-linguistic differences due to word stress location, vowel reduction and consonant clustering.

As we intend to investigate the effect of extrinsic phonetic load of vowel duration on the intensity across prominent syllables, the measure of total amplitude throughout the duration of the syllable nucleus is inappropriate, because within prominent syllables the total amplitude will be higher for extrinsically long vowels (and is not due to prominence). Given that spectral balance is a duration

 ISBN 1-876346-48-5

independent measure, it is a more appropriate measure for this study.

2. METHOD

2.1 SUBJECTS

Data were gathered from 4 Russian and 5 Scottish female middle class speakers, aged between 25 and 45 years old. All speakers were recruited in Edinburgh, Scotland. The Russian speakers had all learnt English at university in Russia, and have been exposed to Scottish for the periods in time ranging from four months to two years.

2.2 STIMULI AND DATA COLLECTION

The data for this study consisted of monosyllabic (CVC type) carrier words. The carrier words contained the vowel /i/ in the syllable nucleus. To control for segmental and voicing influence of the context (1) unvoiced consonants were used in the left context; (2) either unvoiced plosives or voiced fricatives were used in the right context to trigger extrinsic vowel duration differences. The CVC targets were embedded in two types of carrier sentences in four prominent positions. However, for methodological reasons we only took a subset of three positions that will accordingly be referred to as pos 1 – 3 (see Table 1).

Scottish	Russian
It's a [target](pos 1).	Ehto [target] (pos 1).
A [target] is a [target] (pos 2) and nothing but a [target](pos 3).	Tot [target] – ehto [target] (pos 2) i tol'ko tot [target](pos 3).

Table 1 Main type carrier sentences used in the two languages.

Each subject was recorded on a DAT-recorder in a soundproof booth using a condenser boundary microphone with a half-spherical response. The recording volume settings were kept constant. The subject's mouth distance to the microphone was 50 – 60 cm. The subjects were instructed to speak clearly. No specific instructions were provided towards the pitch accent placement in the utterances. Each carrier-sentence was repeated by each subject ten times for each target in an intermittent order with pauses in between.

2.3 DATA ANNOTATION

Utterances containing 160 carrier words in total per subject were digitised at a sampling rate of 11050 Hz and 16-bit quantisation.

For each token the syllable prominence was analysed and labelled. Syllables carrying a pitch accent in a “broad focus” [12] were considered for further analyses, while de-accented and narrow focus syllables were excluded.

Vowel duration was measured after visual inspection of the waveform and the spectrogram of each vowel.

2.3 ACOUSTIC MEASUREMENTS OF SPECTRAL LEVEL AND OVERALL INTENSITY

The spectral level (dB) was calculated in four specified frequency bands from the short term Discrete Time Fourier Transform (DTFT). A Hamming window of 46 ms was used to extract the speech signal samples. The spectral level is 20 times the base-10 logarithm of the measured root-mean-square (RMS) value in a frequency band, relative to the maximum RMS value allowed by 16-bit quantisation.
The four spectral bands (B1 – B4) were defined by the centre frequencies of the four formants individually for each token of /i/. The formant values were measured with the 10th LPC-filter order analysis in PRAAT 4.0. The extracted centre frequencies were then averaged through a steady part of the vowel. Obvious formant analysis errors for target /i/ (e.g. F2 < 1500 Hz) were excluded from the mean calculation. The bandwidths of the spectral bands were fixed as: B1 = mean F1± 150 (Hz) (300 Hz is a reported maximum B1 for female American English speech [11]); B2 =mean F2 ± 300 (Hz); B3 = mean F3 ± 300 (Hz); B4 = mean F4 ± 300 (Hz). For each token, the spectral bands did not overlap.

The RMS power values (dB) were then extracted for each band and means were calculated over the steady part of the vowel. By doing so, we excluded parts of the vowel with possible short-term laryngeal influences from the left and right consonantal contexts. The steady part of the vowel was defined as 25% of the total vowel duration from the onset, and 40% to the offset, the minimum allowed duration of the steady part was set to 25 ms.

Overall intensity was measured and averaged in the same way as in the spectral bands as RMS-power, with the difference that the RMS-power measurements covered all spectrum frequencies.

To normalise for the differences in overall intensity, the spectral band level of each token was expressed as a ratio of the RMS power in a frequency band to the overall RMS power.

2.5 STATISTICAL ANALYSES

To investigate the effect of the right consonantal context on syllable nucleus duration and spectral band level (B1 –B4) in prominent positions we ran a three way analysis of variance (α = .05) with “language” as a between subject factor and “right consonantal context” and “syllable position” as repeated measures.

3. RESULTS

3.1 DURATION

Based on previous studies we expected the extent of the influence of the right consonantal context between the two languages on the duration of /i/ in prominent syllable nuclei to be significantly different. Our results showed that this was indeed the case, with a significant main effect of the right consonantal context in both languages

[F(1,7)=122.189; p<0.01], and a significant interaction between "language" and "right consonantal context" [F(1,7)=76.161; p<0.01]. Table 2 displays differences in extrinsic duration means of the syllable nucleus /i/ across all positions for the two languages.

	Right context	Mean	N	Std.dev.
Scottish	fric +v	209.7	135	40.3
	plos –v	96.2	128	22.5
Russian	fric +v	107.1	64	25.8
	plos –v	90.9	129	15.8

Table 2 Duration (ms) in the two languages for the two right consonantal contexts.

The vowel /i/ is longer before voiced fricatives than before unvoiced plosives in both Scottish and Russian. However, the increase in duration induced by changing manner of articulation and voicing (from unvoiced plosive to voiced fricative) is 118% in Scottish and only 18% in Russian. These numbers confirm previous reports on the extent of extrinsic duration conditioning in Scottish [2] and Russian [3].

3.2 SPECTRAL BAND LEVEL

We hypothesised that the extrinsic phonetic load on the duration of the syllable nucleus /i/ would have a significant effect on the spectral balance in different frequency bands. Statistical analyses revealed that in the lowest frequency band (B1) there were no significant main effects or interactions with the right consonantal context, language or position in utterance on the normalised spectral level. This was not surprising, given that the spectral level of B1 is close to the overall intensity level, and the frequencies were well below the high-frequency range. However, significant effects and interactions were found for all three high-frequency bands (B2–B4). In B2 – B3, we found no significant main effect of either "language" or "right consonantal context", but a significant interaction between the two factors (B2 [F(1,7)=8.628; p<0.05] B3: [F(1,7)=7.127; p<0.05]). In B4, we found a significant main effect of the "right consonantal context" [F(1,7)=10.668; p<0.05] and "language" [F(1,7)=5.791; p<0.05], and no significant interactions. In B2 – B4, we also found a significant main effect of "position" on the spectral level (B2 [F(2,14)=13.609; p<0.01]; B3: [F(2,14)=6.976; p<0.01]; B4: [F(2,14)=6.858; p<0.01]).

Figure 1 shows the means of spectral level ratios in four frequency bands for the two languages as a function of the right consonantal context in all positions. In Scottish, the spectral level in B2 – B4 of /i/ followed by unvoiced plosives is systematically 2 – 5 dB higher than that of the same vowel before voiced fricatives. It also illustrates that for the Russian speakers, the spectral level in B2 – B4 of the vowel before unvoiced plosives is consistently 5–10 dB lower than that of the Scottish counterpart. The spectral level of the Russian vowel before voiced fricatives is in most cases lower than that of the Scottish counterpart. In

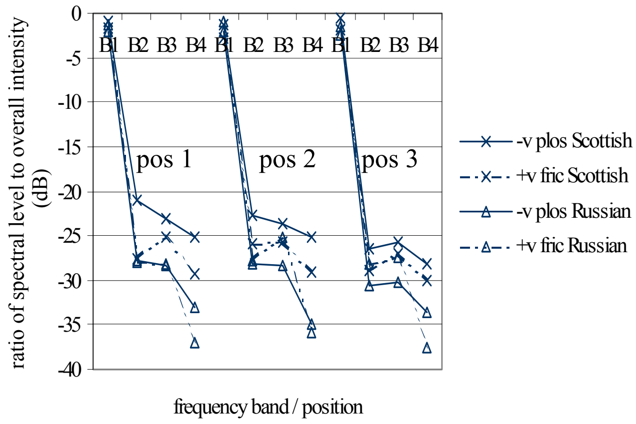

Figure 1 Mean spectral level (dB) in frequency bands in three positions in Scottish and in Russian.

Russian, the influence of the right consonantal context on the spectral levels seems to be less systematic than in Scottish.

It is possible that formant frequencies (F1–F4) of /i/ differ between the languages and between the right consonantal context conditions. As known from the acoustic theory of speech production, a shift in a formant frequency causes spectral level changes at higher frequencies [10]. Therefore, based on the formulas provided in [10] and [8], we calculated how much of the spectral level differences ("raw" in Table 3) can be accounted by formant frequency shifts ("norm" in Table 3). We took the Scottish vowel before unvoiced plosives as a reference and compared its frequencies and resulting levels with the Scottish vowel before voiced fricatives and Russian vowels before unvoiced plosives and voiced fricatives. The results are summarised in Table 3. Although parts of the differences in spectral levels can be explained by the formant frequency shifts, the spectral level ratio differences remain at least 1.2 dB and in most cases well above that value.

Spectral level corrections (dB) in relation to formants of Scottish /i/ before plos –v		B2	B3	B4
1. Scottish /i/ fric +v	raw	4.2	1.9	3.4
	norm	3.9	1.2	2
2. Russian /i/ plos -v	raw	5.8	4.9	7.6
	norm	5.7	4.1	6.6
3. Russian /i/ fric +v	raw	4.9	3.1	10.8
	norm	4.2	2.2	8.4

Table 3. Spectral level ratio differences of /i/ (B2–B3) before and after normalisation for formant frequrency shifts.

4. DISCUSSION

The above results confirm that the extrinsic phonetic load on the duration of syllable nucleus /i/ has a different effect on the spectral balance in all prominent positions in Scottish and Russian. In Scottish, this effect on the spectral

ISBN 1-876346-48-5 © 2003 UAB

level is confined to higher frequencies between 2.5–4.5 kHz (B2–B4), where the short vowel /i/ is produced with a significantly higher spectral level than the long one. This effect can neither be explained by differences in overall intensity nor by formant frequency shifts. It can also not be explained by differences in duration, since there is an opposite effect on spectral levels of short and longer vowels in Russian than in Scottish (at least for B2 – B3). Therefore, we can conclude that the difference in high-frequency levels of the short and long vowel in Scottish is language-specific, and is due to a different glottal source configuration (possibly resulting from greater physiological effort to produce the extrinsically short sound). This relation holds systematically across all prominent positions in this study.

For Russian speakers, the relation of high-frequency spectral levels is less systematic. This is not surprising, because the extrinsic duration differences differences in /i/ are less clearcut in Russian than in Scottish. In general, the mean spectral levels in Russian are well below those in Scottish. This finding confirms the traditional viewpoint of Russian phoneticians, that Russian vowels are characterised by a rather lax articulation even in prominent positions [14]. However, the low spectral levels of Russian in our study may also result from the long-term phonation type employed by our Russian female speakers.

The results further indicate that spectral balance may be a relatively more important acoustic cue to prominence in Scottish than in Russian, since the Scottish speakers utilise spectral balance in a differentiated way for extrinsically short and long vowel /i/ across prominent positions. This conclusion, however, must be treated with caution, as our Scottish speakers may utilise different phonation types in a socio-linguistically marked way (female, middle class). A study into glottal source parameters of the speakers is needed to establish the exact cause of the glottal configuration difference.

It is beyond the aim of the present study to test the perceptive relevance of the spectral level changes in the Scottish short and long vowels. However, it is worthwhile noting, that the high-frequency range, where the high-frequency level ratio appear 1.2 – 4.2 dB higher in the vowel before unvoiced plosives, lays within the region of the highest sensitivity of the human ear (2 – 5 kHz), and the perceptive relevance of spectral tilt has been experimentally confirmed in different studies (e.g. [13][9]).

4. ACKNOWLEDGEMENTS

We would like to thank Peter Rutten for writing a progamme for RMS-power analysis of speech in specified frequency bands and for his useful comments; and Robin Lickley for his help and advice with statistical analyses.

5. REFERENCES

[1] A.J. Aitken, “The Scottish Vowel Length Rule, ” in *So Many People, Longages, and Tongues*, M.L. Benskin, Ed., pp. 131- 157, Edinburgh: Middle English Dialect Project, 1981.

[2] J. Scobbie, N. Hewlett, and A. Turk, “Standard English in Edinburgh and Glasgow: the Scottish vowel length rule revealed,” in *Urban Voices: Accent Studies in the British Isles,* P. Foulkes and G. Docherty, Eds., pp.230 – 245, London: Arnold, 1999.

[3] M. Chen, “Vowel length variation as a function of the voicing of the consonant environment,” *Phonetica*, vol. 22, pp. 129–159, 1970.

[4] N. Svetozarova, “Intonation in Russian,” in *Intonation Systems*, D. Hirst and A. Di Christo, Eds., pp. 261–274. Cambridge: Cambridge University Press, 1998.

[5] L.V. Bondarko, (In Russian) *“A phonetic description of language and a phonological description of speech,*” Leningrad: Leningrad University Press, 1981.

[6] M.E. Beckman, “*Stress and Non-Stress Accent*,” Doordrecht: Foris Publications, 1986.

[7] D.B.Fry, “Duration and intensity as physical correlates of linguistic stress,” in *Journal of Acoustical Society of America*, vol. 27, pp. 765–768, 1997.

[8] A.M.C. Sluijter and V. J. van Heuven, “Spectral balance as an acoustic correlate of linguistic stress,” in *Journal of Acoustical Society of America*, vol. 100, pp. 2471–2485, 1996.

[9] W. N. Campbell, “Loudness, spectral tilt, and perceived prominence in dialogues,” in *Proceedings of ICPhS 95, Stockholm*.

[10] G. Fant, “*Acoustic Theory of Speech Production,*” The Hague/Paris: Mouton, 1970.

[11] H. M. Hanson, “Glottal characteristics of female speakers: Acoustic correlates,” in *Journal of Acoustical Society of America*, vol. 101, pp. 466–481, 1997.

[12] D.R.Ladd, “*Intonational Phonology*,” Cambridge: Cambridge University Press, 1996.

[13] C.Gobl and A. Ní Chasaide,,“Perceptual correlates of source paramenters in breathy voice,” in *Proceedings of ICPhS 99, San Francisco* .

[14] L.V. Bondarko, (In Russian) *“Phonetics of Modern Russian,*” St-Petersburg: St-Petersburg University Press, 1998.